浅析 3D 超写实数字人技术在直播场景中的应用与创新

武海玲 谭景瑜 谷忠伟 支环宇 金 腾
（央视国际网络有限公司，北京 100142）

摘 要：本文以融媒创新节目《两会 C+ 时刻》为例，介绍 3D 超写实数字人应用于复杂直播场景时的总体技术方案、技术难点及应对措施，为基于 3D 超写实数字人开展多元化应用提供借鉴和经验。

关键词：3D 超写实数字人；直播；实时互动；动作捕捉；虚实结合　　**中图分类号**：TN946　　**文献标识码**：A
文章编号：1671-0134（2022）05-014-04　　DOI：10.19483/j.cnki.11-4653/n.2022.05.003
本文著录格式：武海玲，谭景瑜，谷忠伟，支环宇，金腾 . 浅析 3D 超写实数字人技术在直播场景中的应用与创新 [J]. 中国传媒科技，2022（05）：14-17.

1. 背景

2021 年两会，中央广播电视总台央视网推出了两会特别节目“C+ 真探”，总台数字虚拟小编小 C(以下简称“小 C”）首次亮相，以新鲜、独特兼具趣味性的两会报道方式快速出圈。[1]2022 年，央视网继续对数字人技术成果升级并扩大创新应用范围，通过融合数字人系统、实时面捕 / 动捕系统、虚拟演播室系统等多个系统，推出虚拟世界的小 C 与演播室的真人嘉宾、前方两会代表三方同框的两会融媒直播节目《两会 C+ 时刻》，真正让 3D 超写实数字人技术在主流思想的表达中发挥关键作用。



图 1　2022 年《两会 C+ 时刻》直播画面

2. 数字人直播技术方案

本次直播需将数字人小 C、两地嘉宾、虚拟场景、虚拟演播室等多方融合，复杂的直播场景涉及到数据采集、实时渲染、连线播出等多个直播环节，如何实现上下游数据流畅对接实时传输，是本次直播面临的技术难点和挑战。为确保直播画面逼真立体、流畅清晰，在技术层面也需攻关解决数字人面部及动作的实时驱动、与虚拟三维场景等融合实时渲染等多个问题。央视国际网络有限公司通过在 UE 工程融合、灯光融合、数据融合、驱动效果等多个层面进行针对性调优，对每一个直播环节进行逐项优化，实现多个部分的有效联接，最终确保了直播效果和直播安全。



图 2　系统架构图

2.1　数字人面部数据采集驱动

在直播中，小 C 需要基于演播室嘉宾、云连线嘉宾的现场表现，及时做出个性化反馈。数字人的面部表情与口型实时驱动技术是核心也是难点。为提高直播中小 C 面部表情驱动的真实性与灵活性，小 C 面部建模采用了 4D 扫描 + 高保真 3D 数字人像绑定技术，参见图 3。具体分为两个步骤：首先，对静态脸和 FACS 表情进行扫描建模，并以此为基准，对人像面部进行初步 blendshape 绑定；其次，基于前期采集的高精顶点流数据和多阶段顶点流解算算法，在控制亚毫米级误差表情重建的同时，实现准确的跨人像表情语义迁移，并针对实时场景优化人像面部绑定效果。

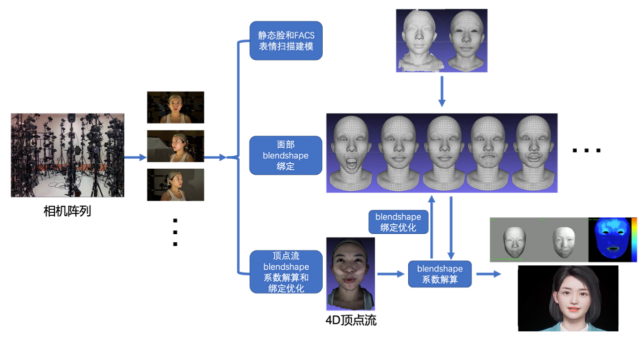


图 3　人像绑定优化与 4D 扫描数据采集、解算

在口型驱动方面，依托前期积累的大量口型数据和

人脸数据标注，通过轻量级的深度神经网络模型进行人像驱动绑定并反复迭代调优，实现端到端的口型还原及表情实时控制，参见图 4。同时，加入词匹配、停顿位置匹配、拼音模糊字匹配等功能，降低匹配错误率，小 C 目前的逐字口型准确率均大于 98.5%，且驱动渲染性能优异、连线延迟低。

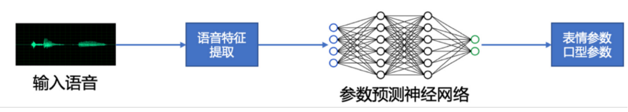


图 4　语音预测口型 / 表情参数

2.2　数字人动作数据采集驱动

在动作驱动上，小 C 采用低时延、高精度、高稳定的专业惯性捕捉方案，把集成了加速度计、陀螺仪和磁力计的 IMU 绑在人体的特定骨骼节点上，通过算法对测量数值进行计算实现动作捕捉。该方案采用先进的微型惯性传感器、生物力学模型以及传感器融合算法，精准捕捉人体上的细微动作，对采集的传感数据经过校准、误差检测和补偿、数据融合后，统一进行三维运动分析。该方案的优势在于实时捕获的空间位移数据不受磁性失真的影响，解决了过去惯性捕捉系统漂移、滑脚以及高度跟踪扭曲等缺陷，在做双脚同时离地、跳跃、爬楼等动作时均不会产生数据漂移现象。同时，输出的动作捕捉数据可靠、纯净、完整，不需要进行后处理即可直接使用。

在实际执行中，借助上述传感器融合技术和动作捕捉引擎，精准采集人的骨骼三维运动数据并分发至虚拟演播室的多个直播机位，平滑连接 MotionBuilder、UE4 等三维引擎系统，成功将稳定、精准又具备个性化特征的运动数据实时映射至小 C 模型（参见图 5），准确还原每个动作，实现在直播中与真人嘉宾的实时流畅互动。

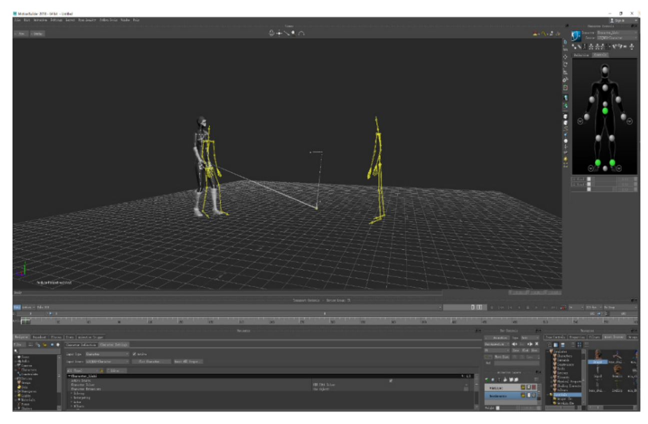


图 5　实时动捕参数配置界面

2.3　虚拟 3D 场景制作

本次直播的主题是“两会 + 科技”，小 C 和现场嘉宾共同置身于一个科技感、未来感超强的虚拟空间进行对话，决定虚拟场景效果的主要因素是灯光。在灯光设计中，使用了基于 PBR 光照模型中的全局光照算法和光学模型来模拟环境整体光线效果，对反照率纹理、法线贴图纹理、金属贴图、粗糙度贴图、环境光遮蔽贴图进行参数化建模（参见图 6），尽量还原现实世界的光照效果。由于场景中的人物和物体对光照的需求不同，在灯光配置时，对场景中的小 C 和虚拟场景利用不同通道进行灯光配置，减少人物光和环境光的相互干扰，并将照向小 C 的灯光与人物的位置关系绑定，避免人物在移动时光源随之移动的问题。灯光的资源消耗是影响直播效果的因素之一。通过对大量光线——物理进行相交测试，进一步提升 3D 空间中路径“采样”的效率，使光线尽可能直接到达观测点或经过单次反射到达观测点，从而减少由于多个灯光重叠而产生的复杂光影对计算资源的消耗。同时，采用 BRDF 对光照反射进行建模，提高渲染效率，有效减少画面噪点。

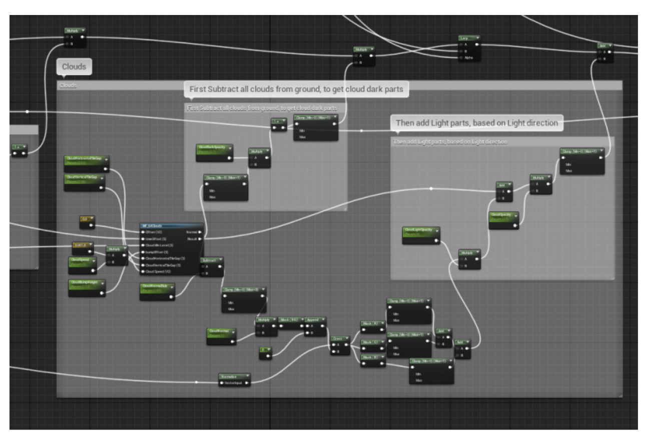


图 6　参数化建模截图

2.4　虚拟演播室融合

虚拟演播室技术将虚拟场景与摄像机现场拍摄的活动图像进行数字化实时合成，使得虚拟元素和画面的三维透视关系完全一致，同步变化，达到逼真、立体的节目效果。本节目使用高精度、可自动跟踪的摄像机，虚拟渲染主机根据摄像机获取的跟踪数据对三维虚拟演播室模型进行位置定位，基于亚像素细节的实时抠像技术，将摄像机信号和虚拟演播室模型进行抠像合成，并将非绿箱部分进行遮罩处理。3DTrackMaker 软件根据跟踪信息制作镜头文件，1：1 还原虚拟物体与人物比例，最终输出逼真丰富的画面。在直播中，还将虚拟场景的相机与现实相机的推拉摇移实时关联驱动，为用户呈现无缝融合的光影效果，实现了高质量的节目播出效果。

2.5　连线播出处理

连线播出系统以超高清切换台为核心，前端可实现摄像机、4K 多通道录放机、手机、笔记本、VR、4G/5G 背包等多种信号源的输入、处理和分发，具备字幕编辑制作、IP 流信号接入、网络推流、多通道多码率录制等

功能。信号输出通过 4K 多通道录放机以及可编辑延时器实现超高清节目录制和超高清信号网络延时直播功能。

在实际直播中，连线播出系统在接入端接入虚拟演播室主备信号、小 C 主备信号、连线嘉宾主备信号等 6 套信号源，由导播台进行镜头切换。小 C、现场嘉宾、连线嘉宾的声音收录后均输入调音台，经 EQ 及音频压缩处理器进行降噪、混音、声画同步处理后，推送至收录设备将处理好的声音与画面合成，最后推流到播出端。在播出端可利用可编辑延时器将直播信号延迟，并推至下游内容分发系统进行多渠道分发播出。

2.6 网络与安全保障

本次直播准备时间紧，直播活动的上下流链条较多，需要一套相对完整且安全的网络环境，服务于虚拟场景、动作捕捉、面部捕捉等多个应用数据分发系统。本次直播网络上采用了扁平化结构，可快速部署、轻维护，并在通讯传输及物理链路上充分考虑备份方案，确保当主路出现问题时可以快速切换到备路，确保安全播出。由于数字人、虚拟场景渲染码率要求高，对网络流量需求较大，故采用高性能网络交换机进行网络连接及数据分发，不同需求的业务通过交换机设置 VLAN 实现网络隔离和多种数据类型交换和传输。参见图 7。

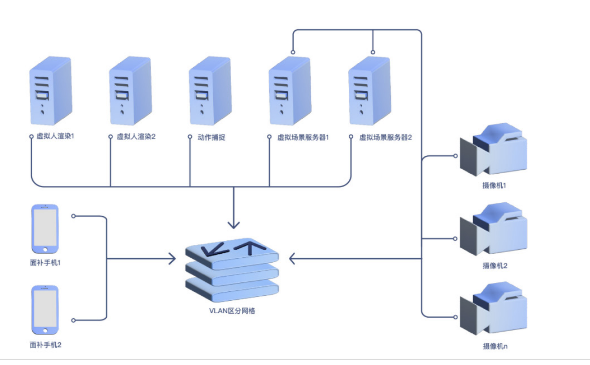


图 7 网络部署示意图

内容安全涉及数据采集、实时渲染、播出画面等多个环节。在数据采集上制定了多台手机面捕备播方案，将多台面捕手机添加至同一局域网中，直播中若主手机出现问题，可自动切换至备用手机。实时渲染层面小 C 特写画面和虚拟演播室均采用主备方案，利用 4 台摄像机和 6 台渲染引擎同时拍摄、渲染，互相作为备份，一旦有一台渲染引擎出现故障，迅速切换其他机位进行节目直播。播出层面则由导播台配置好 6 台渲染机画面和一条备片，提前演练画面切换，在直播中配置专人审核画面效果和画面内容，确保内容安全。

3. 技术难点及应对措施

3.1 低成本无穿戴实时面捕

数字人的面部驱动通常采用捕捉技术采集真人演员的动作和面部表情数据，之后将这些数据迁移合成到虚拟数字人身上，再由动画师手 K 方式处理异常数据，整个流程时间长、成本高。为满足直播低延时、低成本要求，在应用层基于手机自带的 ARKit 框架开发了一套无穿戴式的低成本实时面捕应用。利用手机的景深摄像头采集真人的表情参数和人脸位姿，生成面部 3D 点阵云图，通过表情重定向、位姿校正与映射等处理后，输出 3D 人像的表情参数、人脸位姿（参见图 8），一次采集的数据可以同时发送到四套 UE（主备数字人工程、主备虚拟演播室工程）工程中，实时驱动小 C 做出各种表情，确保多画面输出时小 C 面部效果的一致性，且表情和动作延迟控制在秒级。

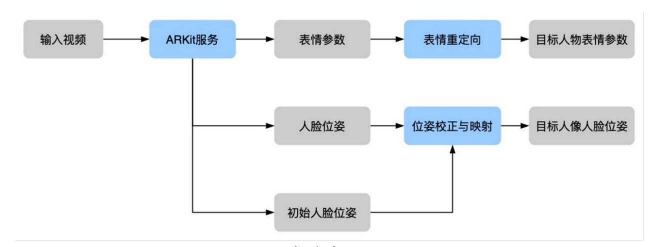


图 8 目标人像表情参数、人脸位姿生成

3.2 直播场景下的多来源 UE 工程融合

现有虚拟演播室系统大多是基于（或兼容）Epic Games 公司的 UE4 系统做深度研发，能够兼容标准 UE4 工程，但在实际使用中因工程个性化开发、多系统协同等原因，往往会出现融合失败现象。本次活动中，在虚拟演播室工程中嵌入小 C 和虚拟场景时，出现了工程迁移不成功的情况，原因是小 C 的 UE 工程既有动画蓝图也有角色蓝图，但虚拟演播室系统一般只兼容动画蓝图。经技术人员多次尝试探索，最终制定了两种可行的解决方案，一是将小 C 工程和虚拟场景融合后，再将虚拟演播室工程迁入融合。二是将小 C 工程与虚拟场景工程融合后独立运行，虚拟演播室系统设置为独立运行模式，最后再将两者合成。最终两套方案在本次直播中均加以应用，实现了虚拟场景、小 C、现场嘉宾在同一画面下的实时动态联动。

3.3 多镜头灯光效果和质量控制

节目播出过程中需呈现多个镜头切换效果，包括小 C 和现场嘉宾的大全景、小 C 特写、嘉宾特写等，为提升整体画面的自然逼真程度，确保用户视觉体验，必须保持多镜头下灯光效果的一致性。UE4 系统拥有完整的虚拟灯光系统，直接影响小 C 和虚拟场景的亮度、色彩以及光影效果。虚拟演播室系统要调整实景环境专业灯光设备对面光、逆光、侧面光、背景光等打光效果，并通过自身的整套灯光控制系统来实现虚实场景灯光融合。两套系统灯光融合效果的好坏，直接影响灯光的输出效果。在联调中，灯光师在最终渲染界面对 UE4 和虚拟演播室灯光的配置参数反复调优，确保画面输出效果统一。此外，过度强调虚拟灯光效果会占用大量的计算资源，影响渲染效果，降低画面质量，所以利用减少灯光数量、

调整灯光位置、强化人物灯光等手段，在有限的计算资源范围内对灯光性能调优，在确保高质量灯光的同时尽量减少计算资源消耗。

3.4 数字人实时渲染画面质量和性能的平衡

虚拟引擎实时渲染的质量控制，本质是分配计算资源、管控性能损耗的过程。小 C 作为 3D 超写实数字人，仅面部面数就接近 10 万面，而面数越多就意味着需要越多的计算资源。同时，虚拟场景的静态及动态灯光、动态阴影、着色器、虚拟演播室的实时抠像合成等都需要大量的计算资源。如何在功能、品质、性能三者中取得平衡，并在直播安全的前提下达到画面的逼真、精细、流畅，这都需要反复测试与验证。经过多天联调测试，在小 C 发型和服装上采用对渲染性能消耗更少的面片头发和较紧身西装，将小 C 独立工程的直播帧率提升至 80FPS 以上；进一步优化虚拟场景和半透明物体控制，去掉较为消耗性能的粒子特效，并将场景灯光聚焦于大全景和小 C 特写，对非核心区域的元素和灯光做减化处理；优化虚拟演播室合成画面，在测试播出画面时发现全景下小 C 出现局部模糊和画面错位问题。问题出在 UE4 系统自身逻辑及系统插件的使用上，通过拉近相机、关掉动态模糊和发质增强、调整灯光设置、调整深度学习超级采样（DLSS）设置等方式，进一步优化了小 C 面部质量和大全景画面效果，参见图 9。通过多种优化举措，最终在直播安全稳定的前提下，实现了画面质量、渲染速度、系统稳定三者的平衡。



图 9 DLSS 高级质量功能打开前后对比图

结语

近期虚拟人呈现井喷式发展，并被广泛应用于媒体场景，而真正将 3D 超写实数字人技术与虚拟演播室技术相结合，并应用于融媒体直播的案例还较少。央视网在《两会 C+ 时刻》节目实践中，综合运用实时面部 / 动作捕捉、实时渲染、深度学习等多种技术，构建了“3D 超写实数字人 + 虚拟演播室 + 直播”虚实结合的直播节目形态，实现了“技术 + 艺术 + 思想”的创新表达。加大 3D 超写实数字人技术应用力度也是业界探索的方向之一，未来将加大 3D 超写实数字人的技术攻关力度，进一步探索将 3D 超写实数字人落地更多应用场景。媒

参考文献

[1] 朱立松，谭志洪，张勇 . 科技成就精彩 智慧赋能媒体——从 2021 年全国两会报道看总台技术的融合创新与发展 [J]. 中国传媒科技，2021（4）：15-20.

作者简介：武海玲（1983-），女，河北承德，中级工程师，研究方向：大数据及人工智能技术的业务应用探索与落地；谭景瑜（1983-），女，广西河池，中级工程师，研究方向：NLP、智能对话、知识图谱、虚拟现实等人工智能技术的研究与应用落地；谷忠伟（1995-），男，安徽亳州，硕士，研究方向：虚拟数字人、XR、大数据、知识图谱等方面的研究与产品研发；支环宇（1988-），男，河北承德，硕士，研究方向：虚拟数字人、智能问答、知识图谱等领域的技术应用；金腾（1984-），男，北京，工程师，研究方向：网络集成和应用。

（责任编辑：陈旭管）

策划语

“智慧广电与广电网络”专题

2021年，国家广播电视总局印发了《广播电视和网络视听“十四五”发展规划》《广播电视和网络视听“十四五”科技发展规划》，明确提出广播电视和网络视听科技工作要围绕六大工程（即舆论引导能力提升工程、新时代精品工程、智慧广电建设工程、视听中国播映工程、安全播出工程、管理优化工程）建设，重点推进落实“发展新业态、培育新视听、建设新平台、重塑新网络、打造新终端、构建新支撑”六项任务。各级广电有线网络公司不断开拓、奋进，着力推动全国一网整合和广电5G建设，深入推进智慧广电工程，加快媒体深度融合步伐，在建设新网络、重塑新内容、培育新业态、构筑新平台、搭建新体系等方面取得了可喜的成就。

本专题系列文章是在中国电子学会有线电视综合信息技术分会、国家广播电视总局科学技术委员会秘书处组织编写的《中国广电有线网络技术年度发展报告（2021）》基础上，对其删改而成。报告起草过程中参阅了部分广电主管部门领导、行业专家、专业调查机构的报告和资料，邹峰、董年初、李戈、彭劲、苗勃、王野秋、陈长伟、聂明杰、马玥等领导和专家参与了讨论、修改，国家广播电视总局广播电视科学研究院给予了大力支持，在此一并致谢！